Customer Segmentation Report

1. Introduction

Customer segmentation is a key strategy for understanding customer behavior and tailoring business strategies accordingly. This report presents the clustering results for customer segmentation based on profile and transaction data using K-Means clustering.

2. Data Overview

The dataset comprises customer profile information (e.g., region, signup date) and transaction data (e.g., total transactions, total amount spent). Feature engineering was performed to derive meaningful attributes for clustering.

3. Clustering Methodology

- **Algorithm Used**: K-Means Clustering
- **Number of Clusters**: 4 (determined using the Elbow Method)
- **Feature Engineering**:
    - Converted `SignupDate` into `SignupDuration` (days since signup).
    - Aggregated transaction data into `TotalTransactions` and `TotalSpent`.
    - Applied One-Hot Encoding for categorical features (`Region`).
    - Standardized numerical features using `StandardScaler`.

4. Clustering Evaluation

- **Davies-Bouldin Index (DBI)**: **1.298**
    - Lower DBI indicates better clustering (lower intra-cluster variance and higher inter-cluster separation). A value of **1.298** suggests a moderately good clustering structure.
- **Silhouette Score**: **0.242**
    - A higher Silhouette Score indicates well-defined clusters. A score of **0.242** suggests moderate separation among clusters.

5. Visualization and Insights

- **Elbow Method**: Used to determine the optimal number of clusters, which was found to be **4**.
- **PCA Visualization**: Principal Component Analysis (PCA) was applied to reduce dimensionality and plot the clusters in a 2D space.
- **Cluster Characteristics**:
    - Some clusters contain customers with high spending and frequent transactions.
    - Others contain customers with low engagement, indicating potential churn risk.

    - Regional distribution varied across clusters, suggesting geographical preferences.

6. Conclusion

The clustering results indicate distinct customer segments based on spending behavior and engagement. These insights can be leveraged for targeted marketing campaigns, personalized offers, and churn prevention strategies.

7. Future Work

- Experiment with alternative clustering methods (e.g., DBSCAN, Agglomerative Clustering).
- Optimize feature selection to improve clustering quality.
- Incorporate additional behavioral data for deeper segmentation insights.